TongueInput: Input Method by Tongue Gestures Using Optical Sensors Embedded in Mouthpiece

Takuma Hashimoto[1] Suzanne Low[1] Koji Fujita[2] Risa Usumi[2] Hiroshi Yanagihara[2]
Chihiro Takahashi[2] Maki Sugimoto[1] and Yuta Sugiura[1†]

[1]Keio University, Yokohama, Japan
(Tel : +81-45-566-1653; E-mail: sugiura@keio.jp)
[2]Tokyo Medical and Dental University, Tokyo, Japan

Abstract: We proposed a system to recognize tongue gestures by mounting a mouthpiece embedded with an array of photo-reflective sensors, to measure the changes in distance between the tongue surface and the back of the upper teeth when the tongue moves. The system utilizes grayscale images of the sensor values to calculate the HOG feature descriptor and to use SVM to recognize the gesture. We conducted two experiments to evaluate the accuracy of the system to estimate 4 tongue positions and 4 tongue gestures, where we obtained a recognition rate of 85.67% for positions and 77.5% for gestures. However, we observed that we can improve the rate by improving the issues we discovered.

Keywords: Tongue Gestures; Photo-Reflective Sensors

1. INTRODUCTION

Input devices using hands and arms are very common. However, using these devices can be challenging for people who have difficulties moving their hands or arms due to neurological or brain disorder. Input devices that helps people with difficulties to play games with a healthy person, encourages them to live a healthier lifestyle.

Some people with neurological disorder like spinal cord lesion have less challenge in moving their eyes, chin, face or tongue, allowing them to operate input devices that track the eye, face, or tongue movements. However, as compare to the jaw movement, input through the eye movement is rather challenging due to the lack of variation and the limited range of movement, while the line of sight may deviate out of the display range.

Our research focuses on the movement of the tongue as using the tongue does not require user to move other parts of their body. Due to wide range of possible motions when using the hands and arms, it is relatively easy to collide onto others. However, as the tongue only operates in the mouth, the concern about obstacles in the surroundings can be eliminated.

The measurement of the tongue motion can be classified into two types; one by using external sensors and another by using intraoral sensors. For external sensors, systems that use magnet [8] or camera [5] are often proposed but both methods require extensive sensors in the surroundings. For intraoral sensors, methods to fix the device on the upper jaw or the front teeth, while using a joystick as the input device have been researched on, where these devices can operate stably and the device size is just right for the mouth. However, it consumes time to fix these devices.

Therefore, we propose an intraoral, easy-to-install sensing system composed of 5 photo reflective sensors embedded on a mouthpiece to detect the movement of the tongue (Figure 1). A photo reflective sensor can irradiate infrared light and measure the reflected light from the object to estimate the distance to the object. A time series data of the sensor values are converted into images, and by clustering the extracted characteristics, we can measure and estimate the gesture of the tongue. As the user is only required to insert the mouthpiece, it does not consume too much time to mount the device.

Figure1. Overview of mouthpiece-type sensor that detects tongue motion by photo-reflective sensors

2. RELATED WORKS

2.1 Estimation of Tongue Motion

Two common methods to measure the tongue movement are to measure using external sensors or sensors inserted internally into the mouth.

In terms of extra sensors, use external sensors. Behnaz Yousefi et al proposed a method to recognize gestures by measuring the change in magnetic field when a riveted magnet is placed on the tongue and 3 magnetic sensors arranged on a headset is placed on both cheeks. However, as this system requires to directly place the magnet on the tongue, it may obstruct the movement of the tongue. SITA developed an interactive tongue training system by measuring the movement of the tongue using computer vision [5]. Goel et al. developed Tongue-in-Cheek, a system that places small Doppler radar unit around the user’s face and measures the Doppler small caused by fine movement of cheeks [4]. Although these divises do not required to be placed around the face.

In terms of intraoral type of sensing system, Saponas et al. embedded 4 optical proximity sensors on to a soft acrylic dental retainer placed on the upper jaw which are fastened by medical grade stainless steel wires to the back molars to recognize 4 tongue gestures: a left swipe, a right swipe, tap up and hold up [3]. Q. Peng and T. F. Budinger developed a system where user can manipulate 5 buttons on a wireless intraoral module using their tongue to send input to the computer [6]. However, placing a device to cover the upper jaw is rather a challenging task to be done by an individual alone.

TonguePoint is a mouthpiece device fitted to individual's upper teeth and hard pallet, containing a 1cm long joystick handle mounted near the roots of the front teeth as a pointing device [1]. Slyper et al. also proposed a similar technique [7]. Although these devices are great operating devices, as the joysticks for operation are relatively small, lots of practices are required. In addition, as it uses occlusal force of the front teeth and it is fixed, it can be quite challenging to be used on a daily basis.

As compare to these researches, our system is completely within the mouth area. Therefore, when the users put on the device, they do not have to concern themselves about hitting their part of the body to other people around them. Our system detects the tongue movement instead of requiring the tongue to move another object, the tongue has more freedom to move, allowing extra input variations.

2.2 Shape measurement around the face using optical sensors

Photo reflective sensors are frequently utilized in facial expression recognition research by using a part of the face as an input to estimate the shape around the face.

An example will be AffectiveWear [2][10], an eyewear device embedded with 8 photo-reflective sensors to detect different facial expressions by measuring the changes in the unevenness of the skin. Similarly, Pupil measures the line of sight and expression simultaneously using photo-reflective sensors focusing on its benefits for remote communication. Utilizing AffectiveWear's idea, AffectiveHMD [9] reflects the user's expression on an avatar in a VR space by embedding photo-reflective sensors arranged on a foldable base onto a head mounted display. AffectiveHMD can also be used to estimate the shape of the mouth.

Photo-reflective sensors can easily be used to create an input interface from the forehead, ears or cheeks etc. DecoTouch [13] and CheekInput [11] are two eyewear devices embedded with multiple photo-reflective sensors to measure the change in wrinkle shape of the skin during touch, to recognize the direction of pulling of the forehead or cheeks or to recognize gestures respectively. EarTouch [12] utilized photo-reflective sensors on an earphone to detect the change in ear shape when the ear is pulled.

All these gesture recognition capabilities can accommodate users to input information into their computer or smartphone by using parts of their body.

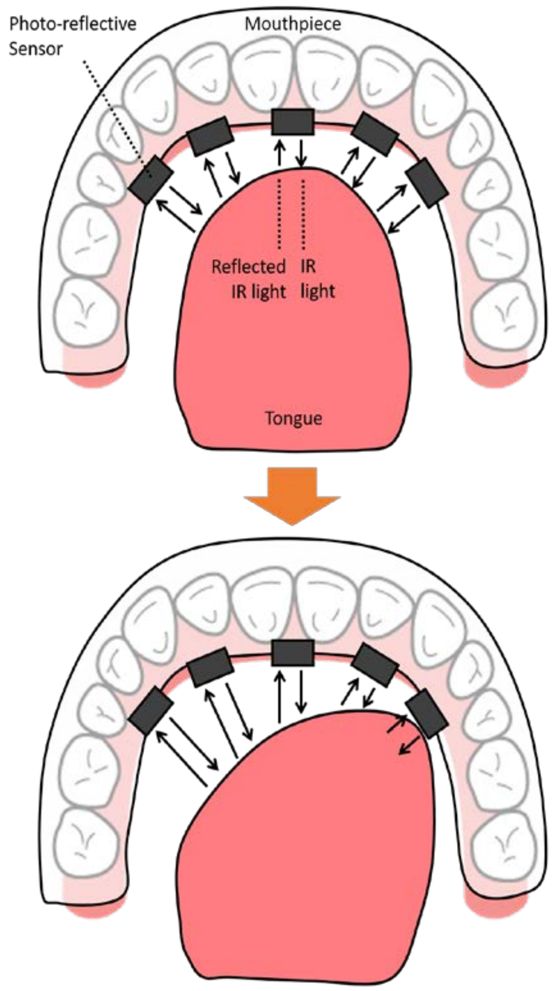

Figure 2. The principle of measuring motion of tongue using photo-reflective sensors

3. SYSTEM IMPLEMENTATION

3.1 Principle

When the tongue moves, the distance between each tooth and the tongue changes. Using this principle, we proposed a system that measures the distance from five locations on the back of the tooth to the tongue using photo reflective sensors. A photo reflective sensor consists of an infrared LED and a phototransistor, with a simple circuit configuration that does not require any special accessory like an amplifier etc. It is commonly used as a sensor to detect the distance of an object. In our system, the sensors are attached to the back of the upper teeth to measure the distance between the sensor and the surface of the tongue.

3.2 System

In this system, we arranged 5 photo-reflective sensors with equal intervals on the inner part of a mouthpiece (Figure 2). We utilized a shape memory type mouthpiece, allowing anyone to use the device. We made a hole on the mouthpiece to place the sensor through the hole. We utilized photo-reflective sensors, connected to an Arduino Pro Mini, 3.3V microcontroller, sending the data to be processed to a laptop PC with Intel Core i7-3770 processor, 8-GB memory using a USB cable (Figure 3).

3.3 Estimation of the Tongue Position

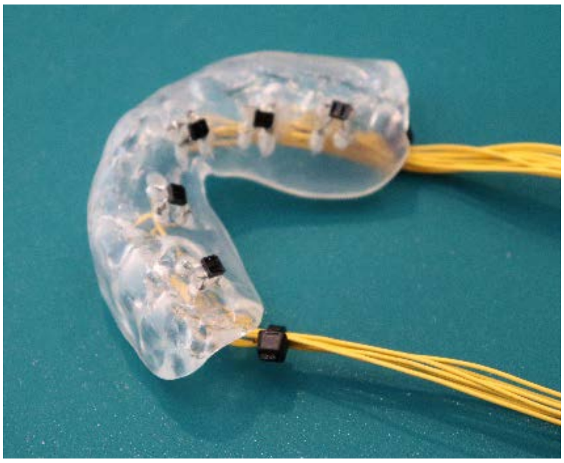

Figure 3. Mouthpiece Type Sensor

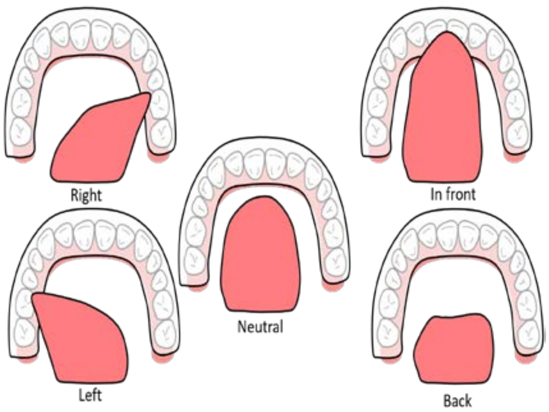

Figure4. Set of tongue position

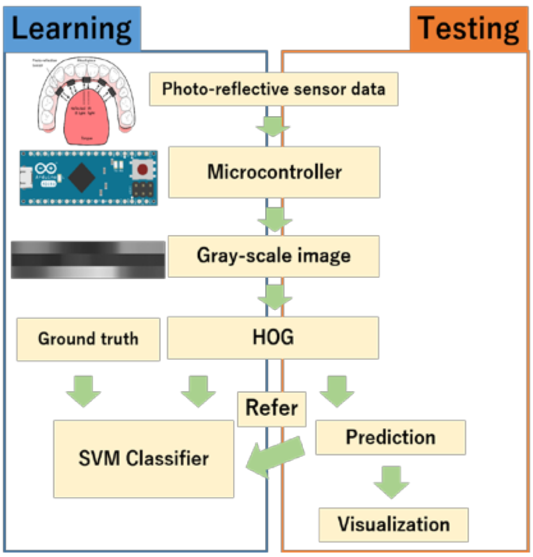

Figure 5. Flow of the estimation of tongue motion

Our system recognizes the tongue position by using the data obtained from the sensors, utilizing the Support Vector Machine (SVM), a commonly used supervised machine learning algorithm for gesture recognition. For the implementation, we used the SVM for processing (PSVM) [Makematics 2012]. The first step is to prepare the direction data. The user wears the device and the system accumulates the learning data by recording the data from the 5 sensors in 4 situations, a hundred times each; when the tip of the tongue moves rightward, leftward, front and back in the mouth (Figure 4). After learning, the system can recognize similar gestures when the gesture is recorded again.

3.4 Estimation of the Tongue Motion

Our system also estimates the tongue motion through the sensors data, by using the SVM algorithm in python3.

The system obtains data from 5 equally distanced sensors, which are integrated to be converted into images. The system then obtains the Histogram of Oriented Gradients (HOG) feature descriptors to be used as the learning and input data for the SVM. HOG feature descriptors are features obtained by generate histgram the gradient direction of the luminance in the local area of the image and are characterized as being robust against geometrical and light changes. The luminance gradient is in 9 directions and the system creates a histogram for each cell with x*x pixels as one cell. It will then normalize the histogram with y*y cells as one block and calculate the feature descriptor. Therefore, the HOG feature descriptor is 9 directions x 2 cells x 19 times = 342 dimensions. The system records the HOG feature of each tongue gesture, and then a SVM learns the HOG features to classify the gestures. We utilized SVM implemented in the scikit-learn library, and set the sampling rate of the system to 100fps.

4. EVALUATION

We conducted two experiments to investigate the accuracy of the tongue position and gesture recognition of the system.

4.1 Evaluation of the Tongue Position

In this experiment, the system will collect the movement of the tip of tongue in four directions as shown in Figure 4, a hundred times each. This is repeated 9 times, equaling to 4500 data collected. We conducted the experiment with the mouth in an open state and started to collect data when the tip of the tongue reached the specified position. Figure 6 illustrates the confusion matrix that shows the experiment results. From the result, we observed that the system could accurately estimate the position of the tongue, showing that the recognition rate is 85.67%.

We also observed that the false recognition rate for Neutral is rather high. The reason is because the participants are not conscious about the normal state of the tongue. In addition, since it's an experiment using a mouthpiece, there is a chance that the participants unconsciously support the mouthpiece with their tongue. There is also a big difference in the movement of the tongue around the top of the upper jaw in Neutral and

Figure 6. Confusion matrix of five-position experiment

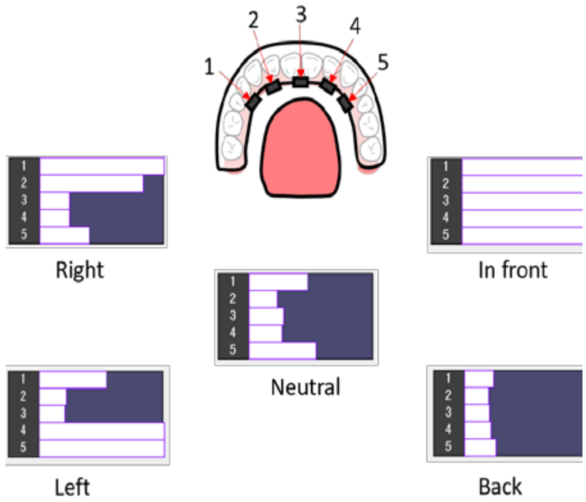

Figure 7. Sensor values of each position

Back that cannot be measured with this sensor array make it challenging to recognize the difference between the two gestures.

We can observe this similarity between the sensor data for both positions in Figure 7.

4.2 Evaluation of the Tongue Motion

In this experiment, we focused on 4 different tongue gestures as shown in Figure 9. The UpperJaw is an action of licking the upper jaw with the tongue, FrontToBack is an action of pulling the tongue out and then rolling it to the back, Tap is moving the tongue up and down at the front teeth position where the tongue touches the teeth intermittently and Circle is an action to lick around the inner part of the mouth using the tip of the tongue. The system acquires 10 images of the imaged version of the sensor value of the 4 gestures and creates a histogram with HOG feature descriptor with 342 dimensions for each sensor value to generate an identifier using SVM.

Similarly, we conducted the experiment with the mouth in an open state. The system obtained the initial sensor data when the tongue is placed at its normal position and then measured the sensor value when the value greatly changes from the initial value. It took 2 seconds to acquire a single sensor value.

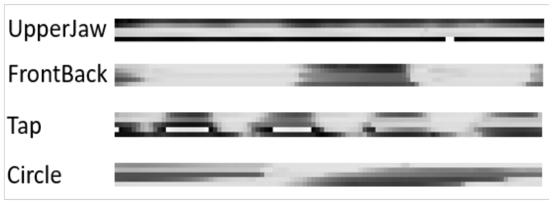

Figure 8. Examples of Grayscale image generated from temporal transition of sensor values of our device for each gesture.

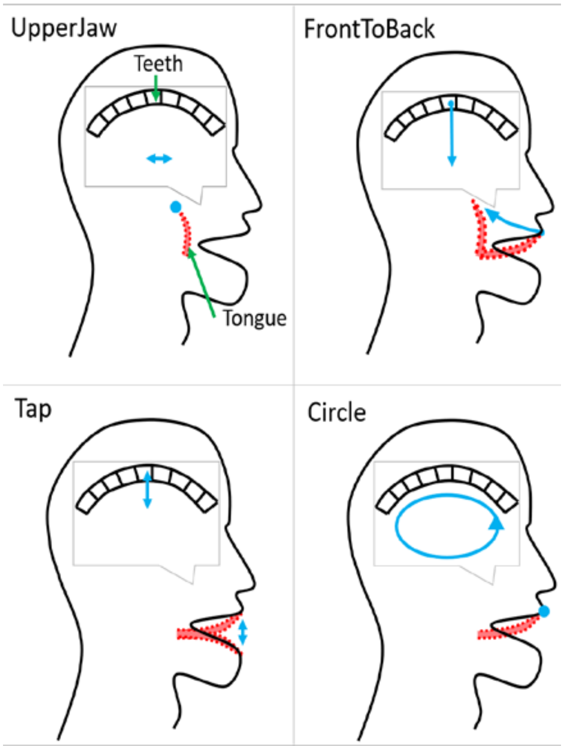

Figure 9. Set of tongue gestures

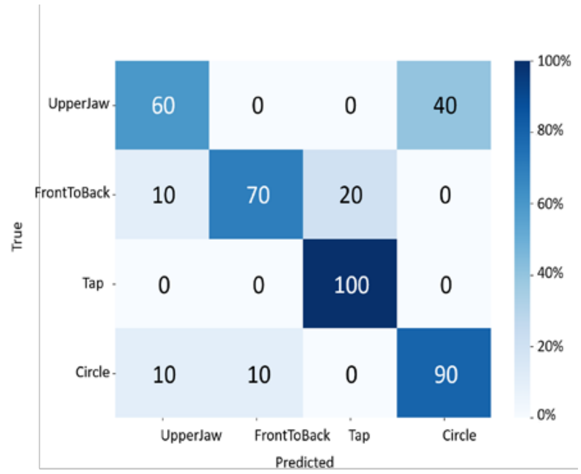

Figure 10.Cofusion matrix of four-gesture experiment

The recognition rate was about 77.5%. Figure 10 illustrates the confusion matrix that shows the experiment results. The results showed that the recognition rate for FrontToBack, Tap and Circle were high, but low for the UpperJaw. From the confusion matrix, we found that half of the UpperJaw cases were mistakenly recognized as Circle as both tongue movements are rather similar to each other. As the 3 gestures other than the UpperJaw have high recognition rate, it shows that the recognition rate will increase if the tongue movement differs from each other.

5. LIMITATION AND FUTURE WORK

Our proposed system has several limitations. First, users are required to practice moving the tongue to follow the right gesture beforehand as it is not common for people to intentionally move their tongue in the suggested gesture movement. Our method is proposed mainly for people with neurological disabilities who have difficulties with voluntary hand and arm movements. Although practice is inevitable, it creates a small barrier for the user to start using the device. Therefore, we would like to redesign simple tongue gestures that do not require any practice beforehand in the future.

The second limitation is that the sensor value for each gesture changes because the shape of each individual's tongue differs. Therefore, it is necessary to create an identifier for each user.

The third limitation is the design of the device. Currently, as the photo-reflector sensors are exposed, the tongue can hit the sensors and this hinders the smooth movement of the tongue. In addition, as each sensor is separately connected to the board, there are about 20 circuits connected in the mouth. Therefore, we will have to redesign the circuit to reduce the amount of wires, to find a material that do not interfere with the attachment of the mouthpiece as well as do not absorb infrared light of the sensor and to examine a better method to fix the sensors.

6. CONCLUSION

We proposed a system to recognize the tongue gestures using 5 photo-reflective sensors arranged in an array fixed to the back of the upper teeth. The system used these sensors to measure the tongue gesture by measuring the distance between the back of the upper teeth and the surface of the tongue. We conducted two experiments to estimate the tongue position and tongue motion. For the first experiment, we found that the system can estimate the position to about 85.67% accurately. For the second experiment, we identified the identification rate of 4 types of gestures. The system used the HOG feature descriptor of generated images of a variation of sensor values that was converted into a histogram as the input to the SVM. From the experiment results, we observed that the system could evaluate tongue gestures that are different from each other. For future works, we would like to improve the design and the required gestures in order to increase the identification accuracy.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP16H01741.

REFERENCES

[1] Chris Salem and Shumin Zhai. 1997. An isometric tongue pointing device. In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (CHI '97). ACM, New York, NY, USA, 538-539. DOI=http://dx.doi.org/10.1145/258549.259021

[2] Katsutoshi Masai, Yuta Sugiura, Masa Ogata, Katsuhiro Suzuki, Fumihiko Nakamura, Sho Shimamura, Kai Kunze, Masahiko Inami, and Maki Sugimoto. 2015. AffectiveWear: toward recognizing facial expression. In ACM SIGGRAPH 2015 Emerging Technologies (SIGGRAPH '15). ACM, New York, NY, USA, Article 4, 1 pages. DOI: https://doi.org/10.1145/2782782.2792495

[3] T. Scott Saponas, Daniel Kelly, Babak A. Parviz, and Desney S. Tan. 2009. Optically sensing tongue gestures for computer input. In Proceedings of the 22nd annual ACM symposium on User interface software and technology (UIST '09). ACM, New York, NY, USA, 177-180. DOI: https://doi.org/10.1145/1622176.1622209

[4] Mayank Goel, Chen Zhao, Ruth Vinisha, and Shwetak N. Patel. 2015. Tongue-in-Cheek: Using Wireless Signals to Enable Non-Intrusive and Flexible Facial Gestures Detection. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). ACM, New York, NY, USA, 255-258. DOI: https://doi.org/10.1145/2702123.2702591

[5] Masato Miyachi, Takashi Kimura, and Takuya Nojima. 2013. A tongue training system for children with down syndrome. In Proceedings of the 26th annual ACM symposium on User interface software and technology (UIST '13). ACM, New York, NY, USA, 373-376. DOI: https://doi.org/10.1145/2501988.2502055

[6] Q. Peng and T. F. Budinger, "ZigBee-based Wireless Intra-oral Control System for Quadriplegic Patients," *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, 2007, pp. 1647-1650.

[7] Ronit Slyper, Jill Lehman, Jodi Forlizzi, and Jessica Hodgins. 2011. A tongue input device for creating conversations. In Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11). ACM, New York, NY, USA, 117-126. DOI: https://doi.org/10.1145/2047196.2047210

[8] B. Yousefi, X. Huo, E. Veledar and M. Ghovanloo, "Quantitative and Comparative Assessment of Learning in a Tongue-Operated Computer Input Device," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 5, pp. 747-757, Sept. 2011. doi: 10.1109/TITB.2011.2158608

[9] K. Suzuki *et al.*, "Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display," *2017 IEEE Virtual Reality (VR)*, Los Angeles, CA, 2017, pp. 177-185. doi: 10.1109/VR.2017.7892245

[10] Y. Lee, K. Masai, K. Kunze, M. Sugimoto and M. Billinghurst, "A Remote Collaboration System with Empathy Glasses," *2016 IEEE International Symposium*

on Mixed and Augmented Reality (ISMAR-Adjunct), Merida, 2016, pp. 342-343.
doi: 10.1109/ISMAR-Adjunct.2016.0112

[11] Koki Yamashita, Takashi Kikuchi, Katsutoshi Masai, Maki Sugimoto, Bruce H. Thomas, and Yuta Sugiura. 2017. CheekInput: turning your cheek into an input surface by embedded optical sensors on a head-mounted display. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17). ACM, New York, NY, USA, Article 19, 8 pages. DOI: https://doi.org/10.1145/3139131.3139146

[12] Takashi Kikuchi, Yuta Sugiura, Katsutoshi Masai, Maki Sugimoto, and Bruce H. Thomas. 2017. EarTouch: turning the ear into an input surface. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17). ACM, New York, NY, USA, Article 27, 6 pages. DOI: https://doi.org/10.1145/3098279.3098538

[13] Koki Yamashita, Yuta Sugiura, Takashi Kikuchi, and Maki Sugimoto. DecoTouch: Turning the Forehead as Input Surface for Head Mounted Display. In Proceedings of the 16th International Conference on Entertainment Computing 2017 (ICEC '17), Springer, 481-484.